Proceedings of the Human Factors & Ergonomics Society's 47th Annual Meeting, October, 2003, pp. 109-113.
(Clearance ASC:03-1666, 19 Jun 2003)

MANUAL VERSUS SPEECH INPUT FOR UNMANNED AERIAL VEHICLE CONTROL STATION OPERATIONS

Mark Draper[1], Gloria Calhoun[1], Heath Ruff[2], David Williamson[1], Timothy Barry[2]

[1]Air Force Research Laboratory
Dayton, OH

[2]Sytronics, Inc.
Dayton, OH

Unmanned aerial vehicle (UAV) control stations feature multiple menu pages with systems accessed by keyboard presses. Use of speech-based input may enable operators to navigate through menus and select options more quickly. This experiment examined the utility of conventional manual input versus speech input for tasks performed by operators of a UAV control station simulator at two levels of mission difficulty. Pilots performed a continuous flight/navigation control task while completing eight different data entry task types with each input modality. Results showed that speech input was significantly better than manual input in terms of task completion time, task accuracy, flight/navigation measures, and pilot ratings. Across tasks, data entry time was reduced by approximately 40% with speech input. Additional research is warranted to confirm that this head-up, hands-free control is still beneficial in operational UAV control station auditory environments and does not conflict with intercom operations and intra-crew communications.

INTRODUCTION

Background

Speech recognition technology enables an operator's speech commands to be used to carry out preset activities. Although speech-based control research has been ongoing for over 25 years, applications have only recently become widespread and accepted by users. This is based on the advancement of automatic speech recognition – significant progress has been made at providing speaker-independent, real-time speech recognition and understanding of naturally spoken utterances with vocabularies of 2000 words and larger (Anderson, 1998). The systems have also matured to the point where they can achieve high recognition rates in noisy environments (Williamson, Barry, and Liggett, 1996).

Application of speech-based input should be pursued to take advantage of this natural and intuitive communication method that allows operators to manage information more efficiently by reducing resource competition, freeing operator's hands, allowing head-up control, and simplifying complex strings of control actions with "voice macros" (and thus reducing error) (Barbato, 1998). These advantages have already been demonstrated in manned aircrew simulations. Speech control improved performance and simplified operations for certain tasks, compared to input made with switches and keyboards (Barbato, 1998). For command and control applications (Theater Air Planning), a speech-input interface improved performance in terms of task completion time over the conventional mouse and keyboard input method (Williamson and Barry, 2000). With speech, the operator simply stated the end menu item and the system brought it up and/or filled in the appropriate information. With conventional manual input, in contrast, the operator had to click through several menu items with extensive "head-down time" and error-prone button selections.

Present Experiment

Command and control stations for unmanned aerial vehicles (UAVs) feature multiple menu pages with systems accessed by numerous keyboard and/or mouse button presses. Thus, the use of speech-based input may also enable UAV operators to navigate through menus and select options more quickly. The present study compared the utility of speech-based input to conventional manual input for data entry tasks performed by operators of a high-fidelity UAV ground control station simulator. Two mission difficulty levels were evaluated as well as different alert modalities. The present paper will report on operator performance with the two input modalities ("Manual" and "Speech") and the impact of mission difficulty. Results pertaining to the alert cue modalities appear elsewhere (Calhoun, Draper, Ruff, Fontejon, and Guilfoos, 2003).

METHOD

Subjects

Ten male Instrument Flight Rules (IFR)-rated pilots served as subjects. Ages ranged from 25 to 48 (mean = 39.8 years). Participants reported normal or corrected-to-normal vision and hearing abilities.

UAV Ground Control Station Simulator

A high-fidelity UAV Air Vehicle Operator (AVO) workstation was used (Figure 1). This station had an upper and a head-level 17” color CRT display, as well as two 10” head-down color displays. The upper CRT displayed an area map (fixed, north up) with overlaid symbology identifying current UAV location, mission waypoints, and current sensor footprint. The head-level CRT (i.e., “camera display”) presented simulated video imagery from cameras mounted on the nose of the UAV. Head-up display (HUD) symbology was overlaid on the AVO’s camera display. The head-down displays presented subsystem and communication information. Visual, auditory, and/or tactile cues alerted operators to abnormal system conditions. The simulation was hosted on six Pentium personal computers. The control sticks were from Measurement Systems Inc. and the throttle assemblies were manufactured in-house to reflect those utilized in current ground control stations. System inputs were made either via a keyboard/trackball (Manual Input) or speech recognizer (Speech Input).

Speech Input

Speech input was achieved with Nuance (Version 8.0.0, Nuance Communications, Inc.), a speaker-independent continuous speech recognition system that supports dynamically extensible grammars. Although Nuance can recognize very large vocabularies (15,000 or more phrases), the vocabulary for the present experiment contained 160 words and short phrases, 70 of which were employed as potential commands for the data entry tasks. To activate the speech recognition system, a “push-to-talk” button was utilized; the operator depressed and held the right side switch on the joystick while speaking the desired voice commands into a microphone (Sennheiser 280-13 Pro headset). Visual feedback of each spoken command was presented on the camera display CRT.

Figure 1. UAV Ground Control Station Simulator.

UAV Operator Tasks

Operators were required to perform a continuous flight/navigation control task while responding to intermittent data entry tasks. Each trial started at 10,000 ft altitude (level off) and involved maneuvering a narrow flight corridor that included either one (low difficulty mission) or three (high difficulty mission) turns. For the flight/navigation tasks, operators were required to minimize deviations from 10,000 ft altitude and 70 knots airspeed while maintaining a position equally distant from the outside boundaries of a narrow flight corridor as indicated on the upper map display. Operators were not allowed to employ the automated “Holds” functions.

The intermittent ‘checklist’ tasks (i.e., series of data entry steps) that an operator had to complete during each trial were representative of operational UAV control tasks. These tasks were classified as Normal Operations, Non-critical Warnings, Critical Warnings, or Information Queries. For Warnings (both Non-critical and Critical) and Information Queries, the operator’s first step was to make a response, confirming detection of an alert cue. After this response, any audio or tactile alerting cue extinguished. For Warnings, the single letter visual cue on the HUD remained as an indicator of the category of warning. For Information Queries, the visual cue contained text indicating what information was to be retrieved. If the operator failed to make a confirmation response within 10 seconds, the cue was extinguished and a “*miss*” was recorded. Operators were allowed 10 seconds to respond to each alert because workload was high when handling multiple tasks.

Assuming the operator detected the alert cue, the remaining procedures were similar across all the data input task types. The required task steps were performed manually or with speech commands, depending on the input modality in effect. Once all the steps were completed, operators made a response denoting task completion. Tasks not completed within experimenter-specified time limits (determined by average manual completion time from pilot study, plus 33%) were scored as “*time-outs*”. For completed and timed-out tasks, the menu automatically returned to the top level to ensure that all tasks started from the same menu page. Table 1 shows each type of data entry task, time limit, required number of button pushes for Manual Input as well as number of speech commands for Speech Input, and the number of tasks per mission difficulty level. Each voice command consisted of a single word or short phrase. Due to the inherent advantages of voice control, many of these functioned as "macros" and effectively replaced numerous sequential button presses.

Design

Each operator flew eight 14-minute experimental trials, four using Manual Input and four using Speech Input, in a within-subjects design. The Input Modality variable was blocked, such that runs were completed with one input modality before runs with the alternate input modality. Within each block of four runs, Alert Modality was blocked and the order of the two runs with each modality, as well as the Mission Difficulty, were counterbalanced across operators and data collection trials. Except for the fact that normal operation tasks occurred at trial start and after each turn, task order was randomized, as well as the time interval between tasks.

Procedures

Operators were first given four hours of training. Practice sessions were conducted for each task separately, then simultaneously, until performance stabilized. Each data entry task was introduced and practiced individually (first with Manual Input, then with Speech Input) prior to flying the entire mission to give pilots the opportunity to train each repeatedly. Prior to each block of four experimental trials, operators completed refresher training with the input modality to be employed next. During all experimental trials, pilots utilized checklist books that detailed the button presses (Manual Input) and commands (Speech Input) required for data entry task completion. Training and experimental trials were completed either in one day or over two consecutive days.

Data Recording

The total time to complete each data entry task was recorded. For "*time-outs*" where operators failed to complete the task before the experimenter-specified time limit, the maximum time limit was utilized. Accuracy measures included the frequency of "*time-outs*", the frequency of tasks completed incorrectly, and the percentage of speech commands correctly recognized. Response time between alert onset and confirmation response was also recorded; tasks where the alert was missed were discarded from the data pool. Root-mean-squared (RMS) error of airspeed, altitude, and path were calculated to measure flight/navigation performance. Subjective ratings were obtained with debriefing questionnaires, including the Modified Cooper Harper rating scale (Wierwille and Casali, 1983).

Table 1. Number of Data Entry Steps to Complete Tasks with Manual and Speech Input.

TASK TYPE (Time Limit in Seconds)	NUMBER OF STEPS		TASK FREQUENCY IN MISSION	
	BUTTON PRESSES	SPEECH COMMANDS	LOW DIFFICULTY	HIGH DIFFICULTY
Normal Operations:			2	4
Level Off Checklist (80)	23	6		
Emergency Waypoint (53)	10	2		
Non-Critical Warnings:			1	3
Datalink Board Overheat (27)	31	3		
GDT Transmitter Overheat (33)	9	3		
Prim/Sec Speeds Differ (53)	22	8		
Critical Warnings:			3	3
Servo Overheat (33)	7	3		
Icing (80)	25	7		
Information Queries (40)	15	4	2	2

RESULTS

Performance across measures was worse in the High Difficulty missions compared to the Low Difficulty missions. Due to space constraints, details will not be presented herein, aside from the fact that there were no significant interactions between Mission Difficulty and Input Modality. The remainder of this section will focus on results pertaining to Manual versus Speech Input.

Task Completion Time

This measure is the time period during which all the required steps for the Normal Operations Tasks, Warnings, and Information Queries were performed (whether accurate or not). Thus, task completion time is a key measure for comparing data entry efficiency with Manual versus Speech Input. Separate Analysis of Variance tests (ANOVAs) were completed on each data entry task type. Results showed that for all task types, task completion time was significantly faster when operators employed Speech Input compared to Manual Input (see Table 2). Average timesavings for data entry tasks ranged from 3.14 seconds (responding to servo overheat warning) to 21.43 seconds (level off checklist). Across tasks, data entry time was reduced by approximately 40% with Speech Input.

Task Completion Accuracy

With regards to the average number of tasks that the operators failed to complete (time-outs) within a trial, an ANOVA showed this was significantly more frequent with Manual Input (mean = 0.95) than with Speech Input (mean = 0.1) ($F(1,9) = 7.974$, $p < 0.05$).

The number of tasks completed incorrectly with Speech Input was less than a third of the number associated with Manual Input. The fact that Speech Input involved fewer steps than Manual Input for all of the tasks was a contributing factor – there were fewer steps to do incorrectly with Speech Input. Additionally, the performance of the speech recognition system was excellent – correct recognition across operators averaged 95.054% (ranging 86.93% to 98.29%).

Response Time to Alerts

For the tasks that included an alert cue, ANOVAs were conducted on the time between alert onset and operator confirmation response (press of space bar or voice command "Confirm") as a function of Input Modality. Results showed that response time was significantly longer for Speech Input than Manual Input (Warnings: $F(1,8) = 16.521$, $p < 0.01$); Information Queries: $F(1,8) = 7.593$, $p < 0.05$). Although statistically significant, the average difference in response times between the two Input Modalities was very short, less than one second.

Table 2. Mean Task Completion Time with Manual and Speech Input.

DATA ENTRY TASK	NUMBER STEPS TO COMPLETE		MEAN TASK COMPLETION TIME (seconds)		
	Manual	**Speech**	**Manual**	**Speech**	**Savings**
Normal Operations					
Level Off Checklist[1]	23	6	56.17	34.74	21.43
Emergency Waypoint[2]	10	2	23.55	13.50	10.05
Non-Critical Warnings					
Datalink Board Overheat[3]	31	3	20.76	11.16	9.60
GDT Transmitter Overheat[4]	9	3	30.21	13.28	16.93
Prim/Sec Speeds Differ[5]	22	8	36.80	25.57	11.23
Critical Warnings					
Servo Overheat[6]	7	3	20.28	17.14	3.14
Icing[7]	25	7	44.12	30.45	13.67
Information Queries[8]	15	4	23.84	11.18	12.66

[1]($F(1,9) = 69.33$, $p < 0.001$) [3]($F(1,7) = 11.534$, $p < 0.05$) [5]($F(1,8) = 51.84$, $p < 0.001$) [7]($F(1,9) = 70.864$, $p < 0.01$)
[2]($F(1,8) = 23.619$, $p < 0.01$) [4]($F(1,7) = 36.554$, $p < 0.01$) [6]($F(1,9) = 10.212$, $p < 0.05$) [8]($F(1,9) = 238.45$, $p < 0.01$)

Flight/Navigation Task

Across all data entry tasks, the RMS airspeed error ($F(1,9) = 3.827$, $p = 0.082$), RMS path error ($F(1,9) = 4.473$, $p = 0.064$) and RMS altitude error (e.g., Level off checklist, $F(1,9) = 8.349$, $p < 0.05$) tended to be less with Speech Input compared to Manual Input.

Subjective Data

The operators rated Speech Input more favorably than Manual Input. On the post-trial data, the operators rated the Manual Input as being more difficult than Speech ($p < 0.01$) and imposing higher workload ($p < 0.01$ on the Modified Cooper Harper Ratings). When asked to compare the two input modalities on the final debriefing questionnaire, operators rated Manual Input worse than Speech Input in terms of interference with flight/navigation task and both speed and accuracy of data entry ($p < 0.01$ for each measure).

DISCUSSION/CONCLUSION

The experimental results were definitive: Speech Input was superior to Manual Input for operators performing in a simulated teleoperated UAV control station environment. Operators' performance was better with Speech Input, both for the flight/navigation task and data entry tasks. Additionally, their subjective ratings indicated Speech was better than Manual Input. The only measure showing an advantage for Manual Input was the time to make a response confirming detection of an alert cue. One contributing factor to this result is that several participants had accuracy problems with the word 'Confirm'. The most typical problem was speaking the word too fast so it sounded like 'Cfirm' or just 'Firm'. Thus, they had to repeat the word several times before successful recognition, inflating the response time. This result may also reflect the time differences between the system acting once the space bar is pressed (Manual Input) compared to the system acting once the push-to-talk button is pressed, the word 'Confirm' is stated, the button is released *and* the speech recognizer has processed the verbal command (Speech Input). Off-line analyses suggest that the system can take up to an additional 1.5 seconds to process a single voice input. Thus, the findings that overall task completion time was better with Speech Input compared to Manual Input for a variety of data entry tasks suggest that the additional "processing time" for each individual voice command is negligible compared to the advantages of Speech Input – head-up, hands-free control that facilitates flight/navigation, improves data entry efficiency through intuitive voice macros, reduces errors, and is a natural, intuitive control input.

Reductions in task completion time might also have been realized by improving how functions are accessed with Manual Input on the menu pages. However, it is anticipated that only slight performance enhancements would result from a different assignment of functions to buttons, etc. This is because the number of functions to be controlled in UAV control stations will remain the same or increase and adding additional buttons is not desirable. Such a solution would also not be as efficient as "voice macros". Moreover, such modifications would not provide the head up, hands-free advantages that the operators preferred with Voice Input. Nevertheless, additional research is needed to confirm that Speech Input is still beneficial in operational auditory environments and does not conflict with intercom operations and intra-crew verbal communications.

REFERENCES

Anderson, T. R. (1998). The Technology of Speech-based Control. In *Alternative Control Technologies: Human Factors Issues*. NATO Research and Technology Organization, RTO-EN-3; pp. 2-1 – 2-9.

Barbato, G. (1998). Integrating Voice Recognition and Automatic Target Cueing to Improve Aircrew-System Collaboration for Air-to-Ground Attack. In *Proceedings of the NATO Research and Technology Organization System Concepts and Integration Symposium*, Ottawa, Canada, pp. 24-1 – 24-11.

Calhoun, G.L., Draper, M.H., Ruff, H.A., Fontejon, J., & Guilfoos, B. (2003). Evaluation of tactile alerts for control station operation. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting.*

Wierwille, W., & Casali, J. (1983). A validated rating scale for global mental workload measurement applications. *Proceedings of the Human Factors Society 27th Annual Meeting*, 129-133.

Williamson, D.T., Barry, T.P., & Liggett, K.K. (1996). Flight Test Results of ITT VRS-1290 in NASA OV-10. *Proceedings of the 15th Annual International Voice Technologies Applications Conference AVIOS '96*, American Voice Input/Output Society, San Jose, CA, 335-345.

Williamson, D.T., & Barry, T.P. (2000). The design and evaluation of a speech interface for generation of Air Tasking Orders. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*, 750-753.